

AD A 049993

DDC FILE COPY

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER (18) AFOSR-78-0044	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) THE LOGIT STATISTIC FOR COMBINING PROBABILITIES - AN OVERVIEW.		5. TYPE OF REPORT & PERIOD COVERED (9) Interim rept.
7. AUTHOR(s) (10) Govind S. Mudholkar E. Olusegun George		6. PERFORMING ORG. REPORT NUMBER
9. PERFORMING ORGANIZATION NAME AND ADDRESS University of Rochester Department of Statistics Rochester, NY 14626		8. CONTRACT OR GRANT NUMBER(s) (15) AFOSR-77-3360
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Rsch/NM Bolling AFB, DC 20332		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS (16) 61102F (17) 2394 (18) A5
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		12. REPORT DATE (11) 1977 (12)
		13. NUMBER OF PAGES 23 (13) 25 p.
		15. SECURITY CLASS. (of this report) UNCLASSIFIED
16. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited.		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number)		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) The problem of combining probabilities occurs principally in combining one-sided independent tests, and in testing the simple hypothesis of goodness-of-fit. The mechanics of using the logit statistic, which is the sum of the logits of the probabilities, in the above two and other applications is desired. Several approaches to studying the classical combination statistics are discussed and the logit statistic is reviewed in their light. Some Monte Carlo experiments involving the logit statistic are summarized.		

DDC
FEB 15 1978
F

AFOSR-TR. 78-0044

THE LOGIT STATISTIC FOR COMBINING PROBABILITIES

- AN OVERVIEW *

by

Govind S. Mudholkar and E. Olusegun George

University of Rochester USA and University of Ife Nigeria

Abstract

The problem of combining probabilities occurs principally in combining one-sided independent tests, and in testing the simple hypothesis of goodness-of-fit. The mechanics of using the logit statistic, which is the sum of the logits of the probabilities, in the above two and other applications is described. Several approaches to studying the classical combination statistics are discussed and the logit statistic is reviewed in their light. Some Monte Carlo experiments involving the logit statistic are summarized.

Approved for public release;
distribution unlimited.

ACCESS	NTIS	DDC	UNANIMOUS	JUSTICE	BY	DISTRIBUTION/PLANT CODES	Dis:	SPECIAL

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH (AFSC)
NOTICE OF TRANSMITTAL TO DDC
This technical report has been reviewed and is
approved for public release IAW AFR 190-12 (7b).
Distribution is unlimited.
A. D. BLOSE
Technical Information Officer

* Research sponsored by the Air Force Office of Scientific Research, Air Force Systems Command, USAF under Grant No. AFOSR-77-3360. The United States Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation hereon.

1. INTRODUCTION AND SUMMARY. A scientific inquiry into any major problem in general consists of several independent investigations separated in time and space and differing in quantitative and possibly qualitative aspects of design. Very often the results of the individual investigations are inconclusive, and it becomes necessary to pool the diverse pieces of evidence. If the problem concerns the truth or falsity of a scientific hypothesis, then in different investigations it may be formulated as different statistical hypotheses, and appropriate tests of significance are applied. The aggregate of these tests, possibly of marginal significance individually, can lead to scientifically decisive conclusions if their results are viewed as a whole. In scientific reporting the most common device used for summarizing results of tests of significance is their P-values. The P-values, also known as the significance probabilities, are simple to interpret marginally and the search for a suitable combination statistic, i.e., a compound of the P-values, on which to predicate an objective judgment about the basic problem is the subject of the theory of combining tests.

Even though many of the classical combination statistics were introduced and are most commonly used for the purpose of combining tests and the greater bulk of literature on their study has grown around this aspect, the simple hypothesis of goodness-of-fit has been an equally strong motivation behind them. Indeed, at an early stage of their development E.S. Pearson (1938) explicitly discussed the dual role of combination statistics, and almost parallel streams of articles on these two applications have developed over the past forty-five years. The well known uniform distribution resulting out of the probability integral transformation is the common

core of the twin applications. Because of it the simple hypothesis of fit in canonical form is the hypothesis of uniformity, and the null distribution of the P-values is uniform. The combination statistics have also been discussed in the solutions to several other problems, e.g. the two sample problem, testing composite hypotheses of fit, and several testing of hypothesis problems in multivariate analysis. The role of the statistics in these solutions stems from the two basic applications mentioned earlier, namely the combination of tests and the goodness-of-fit. However, the study of these statistics, which must perform differently in different solutions has, in the context of these problems, just begun.

Specifically, let T_i , $i = 1, 2, \dots, k$ be k independently distributed statistics, from the k investigations with continuous distribution functions F_{i, θ_i} for testing the respective null hypotheses $H_{0i}: \theta_i = \theta_{i0}$ against respective alternatives $H_{1i}: \theta_i > \theta_{i0}$, $i = 1, 2, \dots, k$. If the large values of T_i are significant, then so are the small values of the P-values $P_i = 1 - F_{i, \theta_{i0}}(T_i)$. On the one hand, the problem of combining the tests is to find a suitable combination statistic $\Psi(P_1, P_2, \dots, P_k)$ for testing the overall null hypothesis $H_0 = \bigcap_{i=1}^k \{H_{0i}: \theta_i = \theta_{i0}\}$, the logical conjunction of H_{0i} , against $H_1: \bigcup_{i=1}^k \{H_{1i}: \theta_i > \theta_{i0}\}$, where $\theta = (\theta_1, \theta_2, \dots, \theta_k)$ and \succ denotes the coordinatewise partial order, viz: $\theta_i \geq \theta_{i0}$ for $i = 1, 2, \dots, k$ with at least one inequality strict. On the other hand, given a sample X_1, X_2, \dots, X_n from a population with distribution function (d.f.) F , the simple hypothesis of goodness-of-fit is $H_0: F = F_0$, where F_0 is a specified d.f.. Under H_0 , $Y_i = F_0(X_i)$, $i = 1, 2, \dots, n$, similar to P_i in the earlier case, are uniformly distributed so that a combination

statistic $\psi(Y_1, Y_2, \dots, Y_n)$ is useful for testing goodness-of-fit.

The best known combination statistics include (i) the earliest proposed $\psi_T = \min_i P_i$ due to H. Tippett, (ii) $\psi_F = \prod_{i=1}^k P_i$ due to R. A. Fisher, (iii) $\psi_P = \prod_{i=1}^k (1 - P_i)$ due to K. Pearson, and relatively new (iv) $\psi_N = \sum \phi^{-1}(1 - P_i)$, where $\phi(\cdot)$ is the standard normal d.f., due to Liptak. These are easy to compute and all have simple null distributions: $-2 \log \psi_F$ and $-2 \log \psi_P$ are χ_{2k}^2 - variables, ψ_N is normally distributed, and ψ_T is distributed as the smallest uniform order statistic. Another statistic of this sort, termed the logit statistic, is considered by George (1977) and George and Mudholkar (1977b). It is the focus of the present paper.

The literature on the theory and methodological and practical applications of the combination statistics is substantial. The theory of combining tests is well summarized in George (1977), and the monograph by Oosterhoff (1969). The applications to the goodness-of-fit problem are described and reasonably referenced in Chapman (1958), Lin (1977), and George (1977). The connection to the two-sample problem may be seen e.g. in Bell, Moser and Thompson (1966); and to the multivariate testing problems in Mudholkar and Subbaiah (1977). This essay reviews recent progress in the study of the logit combination statistics in various applications.

The logit statistic and its null distribution are described in Section 2. In Section 3, the principal approaches to studying the methods for combining independent one-sided tests are surveyed and the logit statistic is reviewed in the light of some of these studies. This section also contains brief illustrative summaries of two simulation experiments conducted with a view to comparing the logit and the classical combination statistics

for combining independent and quasi-independent (i.e. independent only under H_0) tests. The logit statistic is considered in the context of the goodness-of-fit problem in Section 4. In the same section we present some empirical results about the power functions of the variations of a test, for the composite hypothesis of exponentiality, obtained by using Fisher's, Pearson's and the logit statistics. Finally, Section 5 is devoted to miscellaneous remarks including one on the weighted logit statistic.

2. THE LOGIT STATISTIC. Let P_1, P_2, \dots, P_k be k independently distributed P -values, or their analogues in the other applications described in Section 1. Following Berkson (1944) the sum of the log-odds ratios, i.e. the logits, of P_i

$$\psi_L(P_1, P_2, \dots, P_k) = \sum_{i=1}^k \log[P_i/(1 - P_i)] \quad , \quad (2.1)$$

is termed the logit statistic. Under the null hypotheses the P_i are uniform $(0,1)$, the $\log[P_i/(1 - P_i)]$ are distributed according to the logistic law with the d.f.

$$F(z) = [1 + \exp(-z)]^{-1} \quad , \quad (2.2)$$

and consequently ψ_L is distributed according to the k -fold convolution $F_k(\cdot)$ of $F(\cdot)$. It is easy to verify that

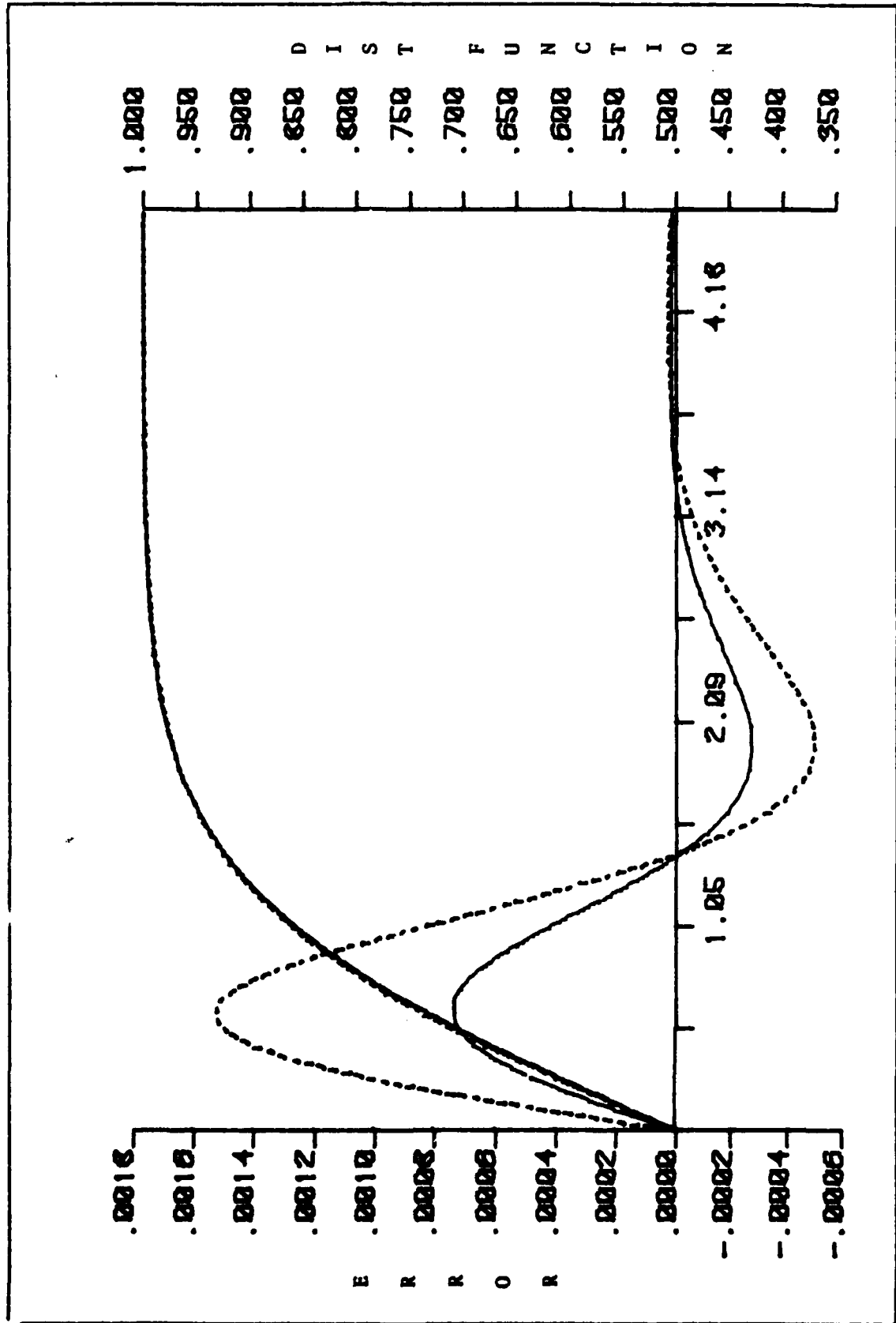
$$1 - F_2(z) = z[e^{-z}/(1 - e^{-z})]^2 + (z - 1)[e^{-z}/(1 - e^{-z})] \quad (2.3)$$

$$\begin{aligned} \text{and} \quad 1 - F_3(z) &= e^{-z}/(1 + e^{-z}) - 2ze^{-z}/(1 + e^{-z})^2 \\ &+ (z^2 + \pi^2)e^{-z}(1 - e^{-z})/2(1 + e^{-z})^3. \end{aligned} \quad (2.4)$$

George and Mudholkar (1977), by inverting the Mittag-Leffler expansion of the characteristic function of ψ_L obtain closed form expressions

FIGURE 1. The Distribution Function* (Right Scale) of the Standardized Logit Statistic, and the Error (Left Scale) in its Approximation by (2.5).

----- k=2 — k=3.



*On the scale the distribution functions, symmetrical about 0, are indistinguishable for k=2 and k=3.

for $F_k(z)$ for general k . However, for practical purposes, the null distribution of the logit statistic admits a simple and very accurate approximation in terms of a student's t -distribution. This approximation is suggested by the fact that the logistic distribution is close enough in shape to the normal distribution to have emerged as a substitute for it in applications such as bioassay. In case of Ψ_L this similarity is enhanced because of the central limit effect. Yet both the logistic distribution and its convolution are heavier tailed than the normal distribution, and may be better approximated by another heavy tailed distribution, namely an appropriate multiple of a student's t -distribution. Specifically, letting t_v be the student's t -variable with v degrees of freedom, it is proposed that

$$\sum_{i=1}^k \log[p_i/(1 - p_i)] \stackrel{\mathcal{L}}{\approx} \pi \sqrt{\frac{k(5k+2)}{3(5k+4)}} t_{5k+4}, \quad (2.5)$$

where $\stackrel{\mathcal{L}}{\approx}$ denotes the equivalence in law, and where the degrees of freedom $(5k+4)$ and the scaling constant $\pi \sqrt{\frac{k(5k+2)}{3(5k+4)}}$ are obtained by equating the variances $(k\pi^2/3)$ and $v/(v-2)$, and the coefficients of kurtosis $(4 \cdot 2 k)$ and $3+6/(v-4)$, of the logit statistic and t_v , respectively. The quality of the approximation (2.5) for $k=2$ and $k=3$ may be seen in

Figure 1, which shows the difference between the d.f. $F_k(z/(k\pi^2/3)^{1/2})$ of the standardized logit statistic $\Psi_L/(k\pi^2/3)^{1/2}$ and the approximation for it. It also displays the d.f.'s for $k = 2, 3$. Even for $k = 2$ the approximation is reasonable, especially in the tails.

3. THE COMBINATION OF TESTS. The importance of the classical combination statistics such as Tippet's Ψ_T , Fisher's Ψ_F and Pearson's Ψ_P lay, at least when they were introduced, in the simplicity of their null

distributions. Since then these have been variously adapted to discrete distributions (e.g. Lancaster (1949), E. S. Pearson (1950), Wallis (1944)) and generalized (e.g. Good (1955)) with some loss of the simplicity. The earliest systematic view of the problem of combining one-sided independent tests described in Section 1 is given by Birnbaum (1954) when he initiated the study of the admissibility properties of the combination methods. Among the workers who have since contributed to the development of the theory of combining tests are Liptak (1958), Lancaster (1961), Schaafsma (1968), van Zwet and Oosterhoff (1967), Oosterhoff (1969), Littel and Folks (1971, 1973), Brown, Cohen and Strawderman (1976) and George (1977). Also noteworthy are the contributions by Davies and Puri (1967), Davies (1969) and others, all stimulated by the needs of Neyman and Scott (1967) for combining rainfall data.

Unlike the common one-parameter problems in statistical theory, the problem of combining one-sided independent tests, a multiparameter problem, does not in general admit solutions which are U.M.P., U.M.P. invariant or U.M.P. unbiased. The theoretical investigations of the subject have therefore centered either on studies of properties such as admissibility, Bayes character, and most stringent character, on derivation of complete class theorems for the tests, or on studies of their asymptotic behavior. In this section we review some properties of the logit statistic ψ_L when used for combining tests in the light of these studies and give some supportive Monte Carlo evidence.

Finite Sample Studies. In order to examine the combination methods for independent one-sided tests theoretically, the problem may also be formulated in terms of the distributions of the P-values instead of, as in the Introduction, in terms of the distributions of the test-statistics

T_i . Both Birnbaum (1954) and Liptak (1958) consider the case in which the only assumption made about the P-values is that the density of each, uniform under the null hypothesis, is nonincreasing in the parameter under the alternative, or equivalently that the family of the distributions of each T_i has monotone likelihood ratio (M.L.R.). For this situation they show that any monotone level α combination test is most powerful against an alternative, where a combination test using a statistic $\Psi(P_1, P_2, \dots, P_k)$ is said to be monotone if rejection of the overall null hypothesis $H_0 = \bigcap_{i=1}^k H_{0i}$ for a vector (P_1, P_2, \dots, P_k) of the P-values implies its rejection for any $(P_1^*, P_2^*, \dots, P_k^*)$ such that $P_i^* \leq P_i$, $i = 1, 2, \dots, k$. All the monotone methods are thus admissible in this mode. Birnbaum, by restricting the model by postulating exponential families for the distributions of T_i , shows that for the admissibility of a combination method, its acceptance region must be convex in the space of T_i , $i = 1, 2, \dots, k$. In this restricted model, which includes tests for normal means when the variances are known but excludes the combination of the t-tests and the analysis of variance tests, the logit method is inadmissible. Liptak on the other hand further examines the subclass of tests based upon statistics of the form $\sum_i \omega_i h(P_i)$ in the original model, where $h(\cdot)$ is monotone and ω_i are nonnegative weights. He demonstrates the Bayes character of such combination tests and shows that they are unbiased if each component test is unbiased. Clearly the logit method, and its weighted version introduced later, shares the admissibility (by virtue of being the most powerful against an alternative), the Bayes character and the unbiasedness with all classical combination tests.

The combination tests have also been studied by assuming that the independent investigations are such that the distribution of the P-values under the alternatives may be assumed to belong to specific parametric families of distributions, as, for example in Yates (1955). For instance, if P_i are independent beta-variables with parameters (a_i, b_i) , then the null hypotheses reduce to $H_{0i}: a_i = b_i = 1, i = 1, 2, \dots, k$ and the most powerful test against simple alternative $\{(a_i, b_i), i = 1, 2, \dots, k\}$ rejects $H_0 = \bigcap_{i=1}^k H_{0i}$ when $\prod_{i=1}^k [P_i^{a_i-1} (1-P_i)^{b_i-1}]$ is large. Thus Fisher's test is U.M.P. against all alternatives satisfying $\{a_1 = a_2 = \dots = a_i < 1, b_1 = b_2 = \dots = b_k = 1\}$ and the logit method is U.M.P. against all alternatives $\{a_1 = a_2 = \dots = a_k < 1, a_i + b_i = 2, i = 1, 2, \dots, k\}$. Lancaster (1961) has developed a more interesting method of evaluating any combination method at a specified alternative by using the series representation of the alternative distribution in terms of the set of functions orthonormal with respect to the null distribution. But such analyses, although very useful locally, are not best suited to overall comparisons of the combination tests.

There are a number of elegant complete class theorems for the multi-parameter problems which are either obtained in the framework of the problem of combining (not necessarily independent) tests or are applicable to it. For example, in terms of the partial order \succ for vectors, defined in the Introduction, Oosterhoff (1969) defines the joint density $f(\underline{t}; \underline{\theta})$, $\underline{t}, \underline{\theta}$ both vectors, as possessing a strict M.L.R. in \underline{t} if for $\underline{\theta} < \underline{\theta}^*$, $[f(\underline{t}; \underline{\theta})/f(\underline{t}; \underline{\theta}^*)]$ is strictly increasing in \underline{t} . If T_i are independent then strict M.L.R. for $f(\underline{t}; \underline{\theta})$ is equivalent to strict M.L.R. for the density of each T_i . For the case of strict M.L.R. if, in addition, $f(\underline{t}, \underline{\theta}) > 0$ for all \underline{t} and the family is dominated by a nonatomic

measure he shows that the class of monotone tests is essentially complete. Brown, Cohen and Strawderman (1976) on the other hand prove a complete class theorem in a more general framework and as an application show that under the above conditions the class of monotone tests is complete. The logit method obviously has membership in such complete classes.

Asymptotic Studies. Oosterhoff (1969) in his monograph describes and uses a number of asymptotic approaches to studying the combination tests. For example, in one, by appealing to the limiting distributions of the test statistics, the problem is reduced to that of combining tests for the means of normal variables with unit variances, and in another the shortcoming of the tests as their levels tend to zero is used as the criterion. However, an effective asymptotic scheme, which is not discussed in the monograph, is Bahadur's (e.g. 1971) method of comparing the exact slopes adapted to the combination problem by Littel and Folks (1971). This method, unlike other asymptotic methods and the traditional approaches described previously, yields measures which describe the operating characteristics of the major combination tests over broad sets of alternatives, and successfully narrows the class of the contenders substantially,

If the null hypothesis is false, then for any fixed alternative the P-value P of any reasonable test converges to zero (exponentially) as the sample size n tends to infinity. The exact slope $C(\theta)$ at an alternative θ , which measures the rate of the decline of P with respect to n , is the a.s. limit of $-(2/n)\log P$, provided that it exists. The computation of the exact slope, in general a nontrivial task, often relies upon a well known technique as summarized, for example, in Theorem 7.2 of Bahadur (1971), which requires the large deviation probability result for the exact null distribution of the test statistic and the a.s.

limit of the test statistic scaled appropriately by a power of n . In the problem of combining the k tests, suppose that P_i , the P -value of the test of $H_{0i}: \theta_i = \theta_{i0}$, is based upon a sample of size n_i , $i = 1, 2, \dots, k$. Let $n = \sum n_i/k$ and denote the d.f. of a combination statistic $\Psi = \Psi(P_1, P_2, \dots, P_k)$, under $H_0 = \bigcap_{i=1}^k H_{0i}$, by $F_n(\cdot)$. Suppose that (i) $(n_i/n) \rightarrow \lambda_i$; as each $n_i \rightarrow \infty$, (ii) $(\Psi/\sqrt{n}) \rightarrow b(\theta)$, a.s. $[\theta]$, where $\theta = (\theta_1, \theta_2, \dots, \theta_k)$, and $0 < b(\theta) < \infty$; and finally (iii) $-(1/n) \log[1 - F_n(\sqrt{n}t)] \rightarrow f(t)$, as $n \rightarrow \infty$, where $0 < f(t) < \infty$, and $f(t)$ is continuous at least for t in the range of $b(\theta)$. Then the exact slope $C(\theta)$ of the statistic Ψ at the alternative θ is given by $C(\theta) = 2 f(b(\theta))$.

In the analyses of the combination statistics composed of the P -values, it is assumed that the exact slope of the component test of H_{0i} at an alternative θ_i , which may as well be regarded as a function of the alternative $\theta = (\theta_1, \theta_2, \dots, \theta_k)$, is $C_i(\theta) = \lim_{n_i \rightarrow \infty} (-2/n_i) \log P_i$ a.s., $i = 1, 2, \dots, k$. With this assumption Littell and Folks (1971) compute the exact slopes of several combination statistics including Ψ_F , Ψ_T and Ψ_N . In order to compute the exact slope of the logit statistic Ψ_L , George and Mudholkar (1977b) use the expressions for the convolution of the logistic distributions obtained earlier (1977a) and show that the large deviation function $f(t)$ for the exact null distribution of Ψ_L is the identity function, i.e. $f(t) = t$. Furthermore, for Ψ_L , $b(\theta) = \sum_{i=1}^k \lambda_i C_i(\theta)/2$. Consequently it is concluded that the exact slope of the logit combination test is

$$C_L(\theta) = \sum_{i=1}^k \lambda_i C_i(\theta), \quad (3.1)$$

the same as the exact slope of Fisher's ψ_F obtained earlier. Littell and Folks later (1973) prove that the exact slope of any monotone combination test is no greater than the exact slope of Fisher's combination test. Clearly, therefore, with respect to the criterion of exact Bahadur A.R.E., the logit combination method is optimal in the class of monotone combination methods.

It is to be emphasized that for obtaining the exact slope as described above, the distribution of the combination statistic is needed only under the overall null hypothesis; under the alternative only the (a.s.) limiting value of the scaled statistic is needed. As a consequence, the method of analysis can be readily applied to the problem of combining tests based on statistics which are independent under the null, but not necessarily under the alternative hypothesis. Using this approach, Mudholkar and Subbaiah (1977) show that if the components of Rao's (1972) test for additional information are combined using Fisher's method then the resulting test is equivalent, in terms of the exact slopes, to the T^2 -test based on all variables. More generally, it is shown that the same equivalence holds between Hotelling's T^2 -test for the problem of testing the significance of the mean of a multivariate normal population, and the Fisher combination of the t-tests in J. Roy's stepdown procedure adapted and investigated by them earlier (1975, 1976) for this problem. Furthermore, in this context, the asymptotic behavior of the logit combination, in this sense, can be shown to be identical to the Fisher combination.

Two Simulation Studies. Two Monte Carlo experiments, which are in progress and are expected to give some indication to the efficacy of the logit method for combining tests as compared with a few classical methods, are now summarized. In the first experiment k independent t-tests based

on samples of size n , for the significance of the means of k normal populations against one-sided alternatives, are combined. In the second experiment Hotelling's T^2 -test for the significance of a mean-vector using a sample of size n from the p -variate normal population is compared with the tests constructed, as in Mudholkar and Subbaiah (1977), by variously combining the p quasi-independent (i.e. independent only under the null hypothesis) t^2 -tests in J. Roy's (1958) stepdown method.

TABLE 1

ESTIMATED POWERS* OF THE COMBINATIONS OF ONE-SIDED
INDEPENDENT T-TESTS, $n = 5$, $\alpha = .05$.

k	Configuration	Test	Noncentrality parameter μ					
			0.0	.2	.4	.6	.8	1.0
2	$\begin{pmatrix} \mu \\ \mu \end{pmatrix}$	Fisher	.056	.125	.280	.478	.678	.852
		Logit	.054	.133	.304	.511	.717	.881
		Liptak	.054	.134	.314	.520	.731	.894
		Pearson	.050	.136	.307	.518	.727	.891
	$\begin{pmatrix} \mu \\ 0 \end{pmatrix}$	Fisher	.055	.081	.145	.191	.291	.395
		Logit	.054	.078	.144	.187	.286	.355
		Liptak	.052	.084	.141	.185	.258	.291
		Pearson	.052	.081	.130	.163	.202	.232
	$\begin{pmatrix} \mu \\ \mu \\ \mu \\ \mu \end{pmatrix}$	Fisher	.054	.180	.445	.716	.920	.985
		Logit	.048	.195	.499	.781	.944	.991
		Liptak	.048	.198	.511	.796	.950	.992
		Pearson	.051	.062	.242	.522	.781	.929
4	$\begin{pmatrix} \mu \\ 0 \\ 0 \\ 0 \end{pmatrix}$	Fisher	.054	.075	.107	.147	.197	.245
		Logit	.048	.077	.108	.140	.180	.213
		Liptak	.048	.074	.108	.137	.169	.182
		Pearson	.051	.072	.094	.100	.120	.134

*Each estimate is based on 3000 samples.

Because of the invariance structures in the problems only normal variables, with unit variance but different means, need to be simulated, for which Marsaglia's (1972) Super-Duper package is used. The P-values and percentiles are obtained using the well known IMSL routines. The power function, which involves k noncentrality parameters in the first

experiment and p in the second, is obtained on a fine grid in the parametric space when $k = 2 = p$. For higher values of k and p , several configurations are included in the simulation but two are always used. In the first, the noncentrality is distributed equally among the alternatives in all the tests, and in the other it is concentrated entirely in only one of the component tests. The above Table I and the following Table 2 give capsule summaries of the currently available results on the power functions being estimated in the two experiments.

TABLE 2
ESTIMATED POWERS* OF T^2 -TEST AND THE COMBINATION TESTS,
 $n = 20$, $\alpha = 0.05$.

p	Configuration	Test	Noncentrality parameter μ					
			0.0	0.2	0.4	0.6	0.8	1.0
2	$\begin{pmatrix} \mu \\ \mu \end{pmatrix}$	T^2	.055	.163	.539	.875	.990	1.00
		Fisher	.056	.163	.548	.885	.992	1.00
		Logit	.054	.163	.546	.888	.993	1.00
		Tippett	.056	.150	.473	.810	.971	1.00
		Pearson	.049	.148	.465	.797	.940	1.00
	$\begin{pmatrix} \mu \\ 0 \end{pmatrix}$	T^2	.049	.104	.285	.581	.836	.963
		Fisher	.048	.105	.285	.578	.830	.961
		Logit	.047	.103	.263	.528	.784	.911
		Tippett	.049	.106	.298	.602	.866	.968
		Pearson	.047	.087	.185	.258	.286	.312
3	$\begin{pmatrix} \mu \\ \mu \\ \mu \end{pmatrix}$	T^2	.049	.192	.629	.948	.998	1.00
		Fisher	.050	.195	.645	.953	.998	1.00
		Logit	.048	.192	.640	.943	.997	1.00
		Tippett	.046	.174	.496	.844	.982	.999
		Pearson	.052	.153	.505	.796	.919	.948
	$\begin{pmatrix} \mu \\ 0 \\ 0 \end{pmatrix}$	T^2	.050	.091	.234	.510	.772	.926
		Fisher	.052	.095	.239	.510	.763	.917
		Logit	.052	.093	.224	.424	.642	.819
		Tippett	.048	.090	.254	.559	.828	.958
		Pearson	.047	.078	.130	.178	.196	.194

*Each estimate is based on 3000 samples.

The consistently poor performance of Pearson's statistic in the two problems is the most striking feature of the two tables. In combining the t-tests, the power of the logit test, at the displayed alternatives, is between those of Fisher's and Liptak's tests. Fisher's test is marginally superior along the coordinate axes, and Liptak's is so along the equi-angular line. However, results not in the table indicate that Liptak's test is considerably inferior at distant points along the coordinate axes without possessing comparable superiority when $\mu_1 = \mu_2 = \dots = \mu_k$. The logit test is a good overall performer for this problem. For Hotelling's problem, when $\mu_1 \neq 0$ and $\mu_2 = \mu_3 = \dots = \mu_p = 0$, Fisher's and T^2 tests are indistinguishable, and both are superior to the logit and slightly inferior to Tippett's test. However, when the noncentrality is equidistributed, i.e. $\mu_1 = \mu_2 = \dots = \mu_p \neq 0$, T^2 , the logit and Fisher's tests are indistinguishable and all three are superior to Tippett's test.

4. THE GOODNESS-OF-FIT PROBLEM. Let X_1, X_2, \dots, X_n be a random sample from a population with the d.f. $F(\cdot)$ and consider the problem of testing a simple goodness of fit hypothesis $H_0: F = F_0$, where F_0 is a given continuous d.f.. Under H_0 the probability integral transforms $Y_i = F_0(X_i)$, $i = 1, 2, \dots, n$, are similar in distribution to the P-values in the problem of combining one-sided tests, i.e. they are uniform (0,1) variables. Hence the combination statistics can also be used for testing H_0 . However the analogy between the two problems does not go very far beyond that, and the results of the studies of the statistics in combining tests are only marginally relevant in the present context. The major factors differentiating the two problems include the following three:

1. Because of the large variety of possible alternative hypotheses, neither small, nor large, or any other specific kind of values of Y_i may necessarily be significant against H_0 . The combination statistics are therefore used for two-sided as well as one-sided tests of fit. 2. In asymptotics, which comprise a substantial portion of the studies, the number k of the test statistics is fixed and $n_i \rightarrow \infty$ in combining tests; whereas in the goodness of fit problem, n , the sample size is allowed to diverge. 3. For the problem of goodness of fit, starting with the work of Kolmogorov (1939), a large class of well known studies and practically applicable alternatives, to the combination statistics, has evolved.

The literature on the study of the combination statistics in the context of the tests of fit is not vast but it is hard to isolate it fully from the truly large body of work on the problem of goodness-of-fit in its generality. Some references to and evaluations of combination statistics as test statistics for testing the hypotheses of fit may be found e.g. in Chapman (1958), Csörgö, Sheshadri and Yalovsky (1975), Hegazy and Green (1975) and Lin (1977). It may be noted that in the problem of combining one-sided tests Fisher's method possesses several optimality properties and is widely used, but Pearson's method is not in general considered to be a contender. In the context of goodness-of-fit, however, neither is superior nor can be eliminated from consideration. The logit statistic which, in a way, is a compromise between the two may be expected to have a good overall performance. At present we are engaged in studies of various aspects and properties of this statistic in testing goodness-of-fit. In this section some of these are briefly outlined.

The null hypothesis $H_0: F = F_0$, in parametric models such as $F = F_0^{\theta+1}$, or $F = 1 - (1 - F_0)^{\theta+1}$, reduces to $H_0: \theta = 0$. By considering various

composite alternatives in terms of sets of θ it is easy to obtain various one-sided and two-sided tests, based upon Fisher's $\Psi_F(Y_1, Y_2, \dots, Y_n)$ and Pearson's $\Psi_P(Y_1, Y_2, \dots, Y_n)$, as the U.M.P. or U.M.P. unbiased tests. In the model $F = [F_0^{\theta+1} + 1 - (1 - F_0)^{\theta+1}]/2$, $0 \leq \theta < 1$, also the null hypothesis is again $H_0: \theta = 0$, and for example, the one-sided test based upon the logit statistic $\Psi_L(Y_1, Y_2, \dots, Y_n)$ is the locally most powerful test. Other models of this variety, their meaning, and the properties of the logit statistic are under investigation.

In terms of the distribution of $Y = F_0(X)$ one such model is

$$\Pr(Y \leq y) = G(y) = \begin{cases} a(y/a)^\theta, & 0 < y \leq a < 1 \\ 1 - (1 - a)[(1 - y)/(1 - a)]^\theta, & 0 < a \leq y < 1. \end{cases}$$

In this model the null hypothesis reduces to $H_0: \theta = 1$, and various two-sided alternatives are considered. These problems have been reduced to simpler problems involving Lehmann-type alternatives by using the transformation $Z = \min [Y/a, (1 - Y)/(1 - a)]$. Chapman (1958) introduces a simple and elegant technique for evaluating maximum and minimum powers of one-sided tests of fit against alternatives at a fixed "distance" from the simple hypothesis. Using this technique he shows that, in terms of the maximum power, Pearson's test is superior to Fisher's test, and also to such conventional tests, as those due to Anderson and Darling, Cramér and von Mises, and Kolmogorov and Smirnov. The price for this superiority is paid in terms of lower minimum power. The logit statistic, a compromise between Fisher's and Pearson's is under evaluation using Chapman's method. More interestingly, as outlined by George (1977), this approach is being extended, by using the Z transform, to two-sided goodness-of-fit tests, and more complex problems involving heaviness of tails.

The asymptotic methods used in the analyses of the goodness-of-fit tests are more traditional and better known, but they do not reduce the class of competing tests to the extent accomplished, using the exact slopes, in the problem of combining tests. Asymptotic analyses of the logit statistic for testing the simple, and some composite hypotheses of goodness-of-fit, using Pitman's, Bahadur's, and Chernoff's (1952) measures of A.R.E. are in progress. We conclude this section by describing a use of the combination statistics for testing the composite hypothesis of exponentiality and an empirical comparison of Fisher's, Pearson's and the logit statistics in this context.

Combination Statistics for Testing Exponentiality. Let X_1, X_2, \dots, X_n be nonnegative i.i.d. random variables and consider the problem of testing the composite hypothesis that their common distribution is exponential. Let $D_{n,i} = (n-i+1)(X_{(i)} - X_{(i-1)})$, $i=1,2,\dots,n$ denote the normalized spacings of the ordered X 's, $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$, $X_{(0)} = 0$. Also let $\delta_r = \sum_{i=1}^r D_{n,i}$, $r=1,2,\dots,n$, and $Z_{r:(n-1)} = \delta_r / \delta_n$, $r=1,2,\dots,n-1$. It is well known that if, and only if, the X 's are exponentially distributed then (i) $D_{n,i}$'s $i=1,2,\dots,n$ are i.i.d. exponentials, and (ii) $Z_{r:(n-1)}$, $r=1,2,\dots,n-1$ are distributed as the $(n-1)$ uniform order statistics.

In view of these characterizations, Csörgö, Sheshadri and Yalovsky (1975) suggest the Pearson statistic $\psi_p = -2 \sum \log Z_{r:n-1}$ for testing exponentiality of the X 's. Obviously, there are several competitors including other combination statistics $\psi_F = -2 \sum \log (1 - Z_{r:n-1})$, $\psi_L = (\psi_F - \psi_p)/2$, and empirical distribution function statistics such as Anderson-Darling A^2 , all to be used in two-sided tests. The following Table 3 contains a small illustrative selection of the results of a

TABLE 3

EMPIRICAL POWERS* OF THE FOUR TESTS OF EXPONENTIALITY;
n = 20, $\alpha = .1$

Alternative	TEST			
	Fisher	Pearson	Logit	Anderson-Darling
χ^2_1	.51	.81	.82	.73
$\chi^2_2 \equiv \text{Exponential}$.100	.100	.102	.103
χ^2_3	.20	.31	.30	.25
χ^2_8	.99	1.00	.93	.91
Lognormal (0,1)	.20	.14	.14	.22
Weibull (.5)	.84	.98	.97	.97
Weibull (2.0)	.97	.97	.98	.97
Beta (2,1)	.46	.28	.47	.40

* Each estimate is based upon 1000 samples.

Monte Carlo experiment performed with a view to comparing these as the tests of exponentiality, over several alternatives. The parameters of the simulation, which uses Marsaglia's (1972) Super-Duper package as a basis for generating various random variables, are $n = 20$, $\alpha = .1$, and each estimate is based upon 1000 samples. More detailed results are given by Lin (1977).

5. REMARKS. The following miscellaneous comments stated in terms of the problem of combining tests are also relevant to the other applications of the combination statistics.

1. Weighted Logit Statistic. Analogous to the weighted version $\Pi P_i^{\omega_i}$ of Fisher's statistic, the weighted logit statistic $\Psi_{L,\omega}$, with weights ω_i , $i=1,2,\dots,k$, is given by

$$\Psi_{L,\omega} = \sum_{i=1}^k \omega_i \log[P_i/(1-P_i)]. \quad (5.1)$$

Under the null hypothesis, $\Psi_{L,\omega}$ may be approximated in law by a scaled t-variable $C.t_v$. The constant C and the d.f. parameter v are determined by (i) equating the variances: $C^2 v/(v-2) = \pi^2 \sum \omega_i^2/3$, and (ii) equating the excesses of kurtosis: $6/(v-4) = (1.2) \sum \omega_i^4/(\sum \omega_i^2)^2$.

The quality of this approximation is expected to be similar to that in Section 2.

2. Selection of the Weights. There is no obvious or simple method known for selecting of the weights ω_i . If there is some clue to the alternative hypotheses then it may indicate some approach to the solution as indicated in Section 3. Otherwise, one may attempt to use the asymptotic argument suggested by Lancaster (1961), or J. Hemelrijk's adaptive approach discussed by Oosterhoff (1969).

3. Discrete P-values. If the P-values are discrete then one method is to use the randomized probability integral transforms, which lead to randomized tests. An alternative, which leads to nonrandomized tests, is to replace the logits $\log[P_i/(1-P_i)]$, using arguments similar to Lancaster (1949), by their conditional expectation, or simpler approximations for them. This method makes such adjustments in the expectation and the variance of the statistic as to render the error due to ignoring discreteness negligible.

4. Two-sided P-values. The Z-transformation mentioned in Section 4, in the context of Chapman's method, is introduced in George (1977) as a one-parameter family, $\min[F_0(T)/\lambda, (1-F_0(T))/(1-\lambda)]$, $0 \leq \lambda \leq 1$, of two-sided P-values. A manuscript discussing the choice of the parameter λ and various applications is in preparation.

5. Exact Slopes and Power. It is now generally recognized that the relationship between the two methods of comparing tests, viz. in terms of the power and the exact slopes, is tenuous. A reasonable empirical approach to investigating the finite sample behavior of tests with equal slopes is to focus on estimating some location parameter the P-values possibly transformed by the logs or logits.

6. Two-Sample Problem. Several aspects of the logit statistic and the Z transformation in the context of the two-sample problem and its one-sample limits, as outlined in Chapter 5 of George (1977), are under study.

ACKNOWLEDGEMENTS. The authors are thankful to Professors C.C. Lin, P. Subbaiah, and Mr. F.C. Pun for discussions and computational assistance and to Professors M.L. Davidson and W.J. Hall for comments and suggestions concerning the manuscript.

REFERENCES

- [1] Bahadur, R.R., "Some Limit Theorems in Statistics," Regional Conference Series in Applied Mathematics, No. 4, SIAM, Philadelphia 1975.
- [2] Birnbaum, A., "Combining independent tests of significance," J. Amer. Statist. Assoc., 49 (1954), 559-575.
- [3] Brown, L.D., Cohen, A. and Strawderman, W.E., "A complete class theorem for strict monotone likelihood ratio with applications", Ann. Statist., 4 (1976), 723-735.
- [4] Chapman, D.G., "A comparative study of several one-sided goodness-of-fit tests," Ann. Math. Statist., 29 (1958), 655-674.
- [5] Chernoff, H., "A measure of asymptotic efficiency for tests of a hypothesis based on sums of observations," Ann. Math. Statist., 23 (1952), 493-507.
- [6] Csörgö M., Sheshadri, V. and Yalovsky, M., "Application of characterizations in the area of goodness-of-fit," Statistical Distributions in Scientific Work, D. Reidel Publishing Company, Vo. 2, 79-90, Boston 1975.
- [7] Davies, R.B., "Beta-Optimal tests and an application to the summary evaluation of experiments," J. Roy. Statist. Soc., Ser. B (1969), 524-538.
- [8] Davies, R.B. and Puri, P.S., "Some techniques of summary evaluations of several independent experiments," Proc. Fifth Berkeley Symposium, Vol. 5, 385-388, Berkeley and Los Angeles, 1967.
- [9] Fisher, R.A., "Statistical methods for research workers," Edinburgh, London: Oliver and Boyd, 4th edition, 1932.
- [10] George, E.O., "Combining Independent One-sided and Two-sided statistical tests - Some Theory and Applications." Unpublished Ph.D. Dissertation, University of Rochester, Rochester, New York, 1977.

- [11] George, E.O. and Mudholkar, G.S., "On the convolution of logistic random variables," Univ. of Rochester Tech. Rep., submitted for publication, 1977(a).
- [12] George, E.O. and Mudholkar, G.S., "The logit method for combining independent tests," Univ. of Rochester Tech. Rep., submitted for publication, 1977(b).
- [13] Good, I.J., "On the weighted combination of significance tests", J. Roy. Statist. Soc., Ser. B 17 (1955), 264-265.
- [14] Hegazy, Y.A.S. and Green, J.R., "Some new goodness-of-fit tests using order statistics," Applied Statistics, 24 (1975), 299-308.
- [15] Lancaster, H.O., "The combination of probabilities arising from data in discrete distribution," Biometrika, 36 (1949), 370-382.
- [16] Lancaster, H.O., "On combination of probabilities: an application of orthonormal functions", Austr. J. Statist., 3 (1961), 20-33.
- [17] Lin, C.C., "Construction and Evaluation of Tests of Fit - Applications of Characterizations and Profiles of Distributions." Unpublished Ph.D. Dissertation, University of Rochester, Rochester, New York, 1977.
- [18] Liptak, T., "On the combination of independent tests," Magyar Tud. Akad. Mat. Kutato Int. Közl., 3 (1958), 171-197.
- [19] Littell, R.C. and Folks, J.L., "Asymptotic optimality of Fisher's method of combining independent tests," J. Amer. Statist. Assoc. 66 (1971), 802-806.
- [20] Littell, R.C. and Folks, J.L., "On the comparison of two methods of combining independent tests," J. Amer. Statist. Assoc., 68 (1973), 223.
- [21] Marsaglia, G., "Random Number Package: Super Duper," McGill University, Montreal, 1972.
- [22] Mudholkar, G.S. and Subbaiah, P., "Unequal precision multiple comparisons for randomized block designs under nonstandard conditions," J. Amer. Statist. Assoc., 71 (1975), 300-307.
- [23] Mudholkar, G.S. and Subbaiah, P., "Testing significance of a mean-vector - A possible alternative to Hotelling's T^2 . Univ. of Rochester Tech. Rep., submitted for publication, 1977.
- [24] Neyman, J. and Scott, E.L., "Some outstanding problem relating to rain modification," Proc. Fifth Berkeley Symp., Vol. 5, 292-325, Berkeley and Los Angeles, 1967.
- [25] Oosterhoff, J., "Combination of One-Sided statistical tests," Mathematic centre Tract 28, The Mathematical Centre, Amsterdam, 1969.

- [26] Pearson, E.S., "The probability integral transformation for testing goodness of fit and combining independent tests of significance, Biometrika, 30 (1938), 134-148.
- [27] Pearson, E.S. "On questions raised by the combination of tests based on discontinuous distribution, Biometrika, 37 (1950), 383-398.
- [28] Pearson, K., "On a method of determining whether a sample of given size n supposed to have been drawn from a parent population having a known probability integral has probably been drawn at random," Biometrika, 25 (1933), 379-410.
- [29] Rao, C.R., Linear Statistical Inference and Its Applications, 2nd ed., New York: John Wiley & Sons, Inc. 1973.
- [30] Roy, J., "Step-down Procedures in Multivariate Analysis", Ann. Math. Statist., 29 (1958), 1177-1187.
- [31] Schaafsma, W., "A comparison of the most stringent and the most stringent somewhere for certain problems with restricted alternative," Ann. Math. Statist., 39 (1968), 531-546.
- [32] Tippett, L.H.C., "The Method of Statistics," First ed., Williams and Norgate, London, 1931.
- [33] Van Zwet, W.R. and Oosterhoff, J., "On the combination of independent test statistics," Ann. Math. Statist., 38 (1967) 659-680.
- [34] Wallis, W.A., "Compounding probabilities from independent significance tests," Econometrica, 10 (1942), 229-248.
- [35] Yates, F., "A note on combination of probabilities test to a set of 2×2 tables," Biometrika, 42, 401-411